The impact of tiredness on virtual reality robotic surgical skills

Alin Adrian Cumpanaș1, Razvan Bardan1, Ovidiu Ferician1, Silviu Constantin Latcu1, Octavian Fulger Lazar2, Ciprian Duta2

1Department of Urology, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania
22nd Department of Surgery, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

Abstract

Introduction: The effect of tiredness has been proved for the surgeons’ musculature performing laparoscopic or robotic procedures (physical stress). Mental stress after robotic surgery has been reported as well. It is still unclear how much the surgical skills are altered and which types of skills are more affected at the final steps of long, complex robotic surgical procedures.

Aim: To evaluate to what extent the surgeon’s skills are influenced by long procedures, using the objective assessment of different surgical skills by a virtual reality robotic simulator.

Material and methods: Fifteen surgeons were asked to perform a continuous 4 h virtual robotic surgical simulator training session. At the beginning of simulator training and at the end of each of the 4 h of training, three exercises of increasing difficulty were selected to be performed in order to assess the surgeons’ skills.

Results: There were statistically significant differences between the initial and final overall scores for all the three exercises, the final outcomes being inferior. The specific metrics for each exercise slightly improved within 1 h from the beginning and thereafter decreased to a statistically significantly inferior value.

Conclusions: The specific metrics on the virtual reality robotic surgical simulator were altered after a 4-hour console training period. Further larger and more complex studies are necessary to evaluate the translation from the simulator to real-life robotic surgery.

Key words: robotic surgery, surgical simulator, surgeon tiredness.

Introduction

Robotic surgery is a fast-growing field, continuously replacing open and laparoscopic procedures in all surgical specialties. The advantages of the robotic system for the surgeons (e.g. 3D visualization, tremor reduction, intuitive motion, Endo Wrist manipulation with 7 degrees of freedom), the surgical outcomes and the patients’ preference are the main driving forces which lead to a steady increase in the number of robotic surgical procedures within the last years [1–4]. In contrast with laparoscopic and open approaches and besides the aforementioned advantages, the robotic surgery approach offers the benefit of ergonomics. Ergonomics in robotic surgery means a more comfortable position sitting at the console, with adjustable viewer height, arm-rest height and foot-pedal location. Taking short breaks during the surgery could possibly improve the surgeons’ further surgical steps. Moreover, when two consoles are available, part of the procedure can be performed by the assistant console surgeon, “passing the ball” between the two console surgeons, minimizing the effects of tiredness on the surgical outcome.

Overcoming the learning curve in robotic surgery is not an easy task, dozens or hundreds of procedures often being necessary to perform in order to
become proficient. Moreover, it seems that open or laparoscopic previous experience does not entirely convert to console surgical skills [5–7]. At the beginning of the individual learning process the length of the robotic procedures can be significantly longer than the corresponding open or laparoscopic approaches. Even for experienced console surgeons, there are robotic procedures which can exceed 4–5 h (e.g. retroperitoneal lymph node dissection, radical cystectomy with intracorporeal neobladder) [8, 9]. In this situation, a question arises: are the surgeons’ skills influenced by tiredness, which could eventually lead to poorer results?

The effect of tiredness has been proved for the surgeons’ musculature (both for laparoscopic and robotic approaches) [10, 11]. There are studies in the literature demonstrating the negative impact of sleep deprivation on surgical performance and in robotic surgical skills [12]. However, it is still unclear how much the surgical skills are altered and which type of skills are more affected at the final steps of long, complex and demanding surgical procedures performed as part of a regular, daily program [13].

Aim

The aim of this study was to evaluate to what extent the surgeon’s skills are influenced by long procedures, using the objective assessment of different surgical skills by a virtual reality robotic simulator.

Material and methods

Fifteen surgeons with previous training exercises on a da Vinci surgical simulator (dVSS-Trainer, Mimic Technologies, Seattle, WA, USA) were enrolled in the study. They were asked to perform a continuous 4 h virtual robotic surgical simulator training session, with a break of 5 min after every hour of training.

The training started in the morning (starting between 8 and 9 a.m.) with a warming-up session of 10 min (when the three monitored exercises/tasks were explained and other types of exercises, on choice, were presented).

At the beginning of the study and at the end of each of the 4 h of training, three exercises of increasing difficulty were selected from the simulator’s menu and were mandatory to be performed by the participants in order to assess the surgeons’ skills. The three exercises were: Peg Board level 1 (consisting in multiple peg transfers on a vertical pad and requiring EndoWrist manipulation, camera movement and navigation, clutching, targeting the objects, grasping the objects and applying the grasping force); Energy Dissection level 2 (which consists in dissections of the small branches of a large vessel using the scissors and the forceps; it requires the skills mentioned for the previous exercise plus dissection tasks, applying monopolar or bipolar energy long enough to seal the vessel and switching between the monopolar and bipolar energy sources); and Suture Sponge level 3, the most complex of them (consisting in sutures on a virtual sponge, through specific dots and with a specific robotic arm, forehand and backhand sutures with the right hand and with the left hand, requiring EndoWrist manipulation, camera movement and navigation, clutching, targeting the objects, grasping the objects, applying a grasping force, needleholder control and needleholder driving). The results of these exercises were recorded and analyzed. Apart from these exercises, in every hour, the participants had to perform at their choice other types of exercises from the simulator menu which were neither registered nor assessed.

The participants were asked to postpone the completion of the study exercises if they had a night shift the night before the study, in order to avoid potential biases which could arise (e.g. accumulative tiredness due to surgical procedures during the night, sleep-deprivation consequences).

The study was conducted according to the national requirements and fulfilled the local ethical committee approval criteria. Consent was obtained from all the participants of the study, being explained the study objectives, methods and that all data were to be analyzed anonymously. The study was carried in accordance with Good Clinical Practice Guidelines implemented in the World Medical Association Declaration of Helsinki.

Statistical analysis

Statistical analysis was performed using the Kruskal-Wallis test. Data were displayed as mean values ± standard deviation. Mean and standard deviations were expressed as continuous data. A p-value < 0.05 was considered significant.

Results

The mean age of the study group was 34 years (range: 27–53 years). All the participants in the
The impact of tiredness on virtual reality robotic surgical skills

The overall score for the three exercises, comparing the results at the beginning of the training session (initial score) with those after 4 h of virtual reality simulator practice (final score), are presented in Table I. One can see that, for overall scores, there is a statistically significant difference between initial and final scores for all the three exercises.

The specific metrics for the first exercise (Peg board level 1), detailed by hours, are presented in Table II.

For the easiest exercise (Peg board, level 1), the detailed analysis reveals the steady improvement of most of the specific metrics (time to complete, instrument collision, excessive force applied to instruments, instruments out of view) at 1 h and 2 h. Thereafter, the specific metrics have a tendency to slightly alter.

The results for exercise 2 (including more complex tasks) are presented in Table III.

The results for exercise # 2 (Energy dissection level 2) show the same pattern as observed in exercise # 1: the specific metrics improve within the first

### Table I. Overall initial and final scores for the three exercises

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Mean value ± SD</th>
<th>Mean value ± SD</th>
<th>Statistical significance, p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise 1 (Peg board – level 1)</td>
<td>84 ±19</td>
<td>77 ±21</td>
<td>0.039</td>
</tr>
<tr>
<td>Exercise 2 (Energy dissection – level 2)</td>
<td>75 ±15</td>
<td>68 ±19</td>
<td>0.038</td>
</tr>
<tr>
<td>Exercise 3 (Suture Sponge – level 3)</td>
<td>66 ±19</td>
<td>56 ±22</td>
<td>0.040</td>
</tr>
</tbody>
</table>

### Table II. Specific metrics, by hours, for exercise 1 (Peg Board level 1)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial results Mean value ± SD</th>
<th>After 1 h of activity Mean value ± SD</th>
<th>After 2 h of activity Mean value ± SD</th>
<th>After 3 h of activity Mean value ± SD</th>
<th>Final results (after 4 h of activity) Mean value ± SD</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete [s]</td>
<td>92 ±23</td>
<td>91 ±24</td>
<td>90 ±22</td>
<td>94 ±24</td>
<td>97 ±25</td>
<td>Initial vs. 2 h, p = 0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.028</td>
</tr>
<tr>
<td>Instrument collision</td>
<td>0.88 ±30</td>
<td>0.89 ±28</td>
<td>0.85 ±32</td>
<td>0.89 ±30</td>
<td>0.95 ±32</td>
<td>Initial vs. 2 h, p = 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.041</td>
</tr>
<tr>
<td>Excessive force applied to instruments</td>
<td>0.07 ±0.01</td>
<td>0.06 ±0.01</td>
<td>0.05 ±0.02</td>
<td>0.09 ±0.01</td>
<td>0.1 ±0.01</td>
<td>Initial vs. 2 h, p = 0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.041</td>
</tr>
<tr>
<td>Instruments out of view</td>
<td>0.32 ±0.05</td>
<td>0.29 ±0.04</td>
<td>0.27 ±0.06</td>
<td>0.34 ±0.05</td>
<td>0.35 ±0.06</td>
<td>Initial vs. 2 h, p = 0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.045</td>
</tr>
<tr>
<td>Economy of motion [cm]</td>
<td>145 ±38</td>
<td>134 ±42</td>
<td>135 ±46</td>
<td>146 ±44</td>
<td>156 ±47</td>
<td>Initial vs. 2 h, p = 0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.045</td>
</tr>
<tr>
<td>Drops</td>
<td>0.20 ±0.04</td>
<td>0.21 ±0.06</td>
<td>0.18 ±0.06</td>
<td>0.22 ±0.07</td>
<td>0.37 ±0.05</td>
<td>Initial vs. 2 h, p = 0.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.038</td>
</tr>
<tr>
<td>Master work-space</td>
<td>9 ±2</td>
<td>8 ±1</td>
<td>8 ±2</td>
<td>10 ±3</td>
<td>12 ±3</td>
<td>Initial vs. 2 h, p = 0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 h vs. final, p = 0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Initial vs. final, p = 0.038</td>
</tr>
</tbody>
</table>

*Bold* values for statistically significant values; NS – no statistical significance.
2 h, after which the subjects’ performance worsens statistically significantly, with two exceptions – misapplied energy time and instrument collision.

Table IV presents the results for the most complex exercise (Suture Sponge level 3).

One can observe that there is the same trend in altering the subjects’ performance on specific metrics for the most demanding exercise. Besides the statistical significance, the difference in absolute values between the initial and the final results of excessive force and mastering the workplace are much higher compared with the previous two exercises. Meanwhile, the instrument collision specific metrics remain unmodified, being the only one not influenced by tiredness when subjects were evaluated on their performance on the most complex exercise.

Discussion

With the tremendous development of robotic surgery, more complex surgical procedures are reported to be feasible by this approach. This means longer, skill-demanding and challenging procedures, even for experienced console surgeons. For open surgery, there are high-volume centers which prefer, in the case of long, complex surgical procedures, to use two surgical teams (e.g. for radical cystectomy with an orthotopic neobladder in the University Medical Center in Mansoura, Egypt there is one team doing the cystectomy and lymph node dissection, followed by the team which performs the orthotopic neobladder). This way, both teams are focused on specific, shorter tasks, avoiding the approach of the most sensitive and skill demanding steps (e.g. reconstruction of the bladder) with tired surgeons.

The use of a virtual robotic surgical skills simulator (dVSS-Trainer) offers, beside the training value, the opportunity to objectively assess the surgical skills, their improvement or alteration with time. By choosing different types of exercises, with different levels of difficulty, one can evaluate with reliable accuracy the ex-vivo surgical skills [14, 15]. Among the long list of exercises offered by the dVSS-Trainer platform, the three exercises we chose for this study (requiring two-handed maneuvers, camera movements, cautery use, etc.) have been described as in-
including reliable and balanced tasks in order to offer a close-to-reality evaluation of the surgeons’ skills and do not depend on the previous surgical experience [6, 16].

A study performed on more than 400 surgeons reported that 56% of them described physical symptoms or discomfort (neck stiffness, finger and eye fatigue) after robotic surgical procedures [10]. Does it translates into poorer surgical skills at the end of the procedure? A recent study comparing the levels of mental workload using the PFS (Piper Fatigue Scale) prior to the surgery and after the surgery revealed significantly increased levels after surgery for both laparoscopic and robotic surgeons [17]. Until now, most of the published studies have evaluated the ergonomic aspects of the robotic surgery and/or the physical and mental stress during the robotic surgery [11, 18, 19]. To our knowledge, this is the first study to evaluate the impact of longer robotic procedures on surgeons’ skill using the virtual reality simulator.

How the surgeons’ fatigue would affect specific tasks requiring attention and dexterity is still unclear. Functional brain magnetic resonance imaging (MRI) studies on sleep-deprived people showed the decrease of the signals from the dorsolateral prefrontal cortex, which has an important role in performing high-demanding, sustained attention tasks [20]. There are studies in the literature showing that starting non-emergent complex surgical procedures late in the day is associated with 2 times higher absolute and risk-adjusted mortality [21].

Besides the overall score for the three exercises at the beginning of the study and after 4 h (which are statistically significant for all of the exercises), it is interesting to analyze the specific metrics of the three exercises. Most of them change according to the same pattern: improving within the first hour, showing a statistically significant deterioration at 2 h and continuing to worsen until the end of the study period. No statistically significant changes were noted between increments of 1 h for each of the specific metrics of the three exercises.

The specific metrics that did not alter statistically significantly during the study were: instrument collision (initial vs. 2 h, 2 vs. 4 h and initial vs. final),

Table IV. Specific metrics, by hours, for exercise 3 (Suture Sponge level 3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial results Mean value ± SD</th>
<th>After 1 h of activity Mean value ± SD</th>
<th>After 2 h of activity Mean value ± SD</th>
<th>After 3 h of activity Mean value ± SD</th>
<th>Final results (after 4 h of activity) Mean value ± SD</th>
<th>Statistical significance</th>
</tr>
</thead>
</table>
| Time to complete [s]        | 250 ±30                         | 240 ±35                               | 231 ±32                               | 245 ±34                               | 270 ±31                                                | Initial vs. 2 h, $p = 0.032$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.036$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.042$
| Economy of motion [cm]      | 344 ±68                         | 325 ±64                               | 322 ±66                               | 356 ±64                               | 370 ±62                                                | Initial vs. 2 h, $p = 0.034$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.037$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.032$
| Instrument collision        | 2.5 ±0.5                        | 2.3 ±0.7                              | 2.3 ±0.5                              | 2.6 ±0.5                              | 2.8 ±0.4                                                | Initial vs. 2 h, $p = 0.07 – NS$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.07 – NS$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.06 – NS$
| Excessive force             | 0.3 ±0.02                       | 0.2 ±0.04                             | 0.3 ±0.05                             | 0.4 ±0.04                             | 0.6 ±0.02                                                | Initial vs. 2 h, $p = 0.2 – NS$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.027$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.032$
| Instruments out of view     | 0.81 ±0.15                      | 0.78 ±0.2                             | 0.71 ±0.2                             | 1.35 ±0.1                             | 1.42 ±0.2                                                | Initial vs. 2 h, $p = 0.034$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.026$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.032$
| Drops                       | 0.2 ±0.02                       | 0.2 ±0.01                             | 0.1 ±0.02                             | 0.3 ±0.03                             | 0.6 ±0.03                                                | Initial vs. 2 h, $p = 0.028$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.032$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.042$
| Master workspace            | 7.3 ±1.3                        | 6.8 ±1.2                              | 6.5 ±1.2                              | 7.4 ±1.3                              | 9.3 ±1.1                                                | Initial vs. 2 h, $p = 0.038$
|                            |                                 |                                       |                                       |                                       | 2 h vs. final, $p = 0.027$
|                            |                                 |                                       |                                       |                                       | Initial vs. final, $p = 0.032$

*Bold values for statistically significant values; NS – no statistical significance.*
instruments out of view (initial vs. 2 h) and misapplied energy time (initial vs. 2 h, 2 vs. 4 h and initial vs. final) for exercise 2 and instrument collision (initial vs. 2 h, 2 vs. 4 h and initial vs. final) and excessive force (initial vs. 2 h) for exercise 3. It seems that the instrument collision and misapplied energy time are more frequently not affected by the long lasting simulator practice. Meanwhile, instruments out of view (exercise #2) and excessive force applied to instruments (exercise #3) are not statistically significantly worsened after 2 h of practice but in the end are statistically significantly poorer than at the beginning. This can be explained by the role of working memory. The working memory was described in sleep-deprivation studies as keeping task-relevant information for a few seconds in order to be used for routine ongoing activities, and it has been proved that it is not always affected by sleep deprivation/tiredness [22, 23].

One can see that specific metrics such as time to complete, economy of motion, drops and master workspace suffer statistically significant alteration after 2 h of console practice, altering the overall score for all the three exercises. We can say that these are the most specific metrics prone to change as the surgeon gets tired.

The strengths of this study are that the study group is a homogeneous one, with surgeons with open and laparoscopic experience, having experience in console training exercises only, without robotic surgical experience. By this, we aimed to avoid biases due to different robotic surgical experience on simulator exercises. Another strength was the study’s design including the choice of the exercises (type, level of difficulty) and the simple and clear console program to obtain reliable results and to avoid biases. Between the hourly evaluations, the participants were allowed to perform training exercises at their own choice, to avoid the boredom described to occur when the same exercises are repeated for a long period of time [24]. The study limitations are represented by the relatively small number of participants and the study length, limited to 4 h. A larger study will be started, including longer simulator periods and an increased number of tasks and situations to be evaluated.

Having said this, what could be the clinical implications? Being a simulator study, it is questionable to what degree the obtained results could be translated into real-life robotic surgery, but it is reasonable to consider so, as long as emerging data from the last years report the surgeons’ fatigue as an important factor affecting the mortality and complication rates [25, 26]. Whether further, larger studies would confirm our results, one should take into consideration the use of the major advantages of robotic surgery which could combat the effect of fatigue (e.g., two-console surgery, changing the console surgeon or taking breaks to combat fatigue).

Conclusions

The specific metrics on the virtual reality robotic surgical simulator are altered after a 4-hour console training period. After the first hour of training (when the performance improves), the surgeons’ performance slightly decreases to a statistically significant difference at the end of the study. Further larger and more complex studies are necessary to evaluate the translation from the simulator to real-life surgery.

Conflict of interest

The authors declare no conflict of interest.

References

The impact of tiredness on virtual reality robotic surgical skills


Received: 19.01.2020, accepted: 27.01.2020.